

***In-silico* based development, identification and characterization of EST based SSR from Cinnamon**

Raman Bhardwaj, Raghvendra Saxena, Rajesh Singh Tomar, Raghvendra Kumar Mishra*

Amity Institute of Biotechnology, Amity University Madhay Pradesh Gwalior (M.P.)- India

Corresponding Author:

rkmishra@gwa.amity.edu

ABSTRACT

Background: The discovery of genes and to construct genome mapping simple sequence repeats (SSR) markers plays very crucial role which were derived from the expressed sequence tags (ESTs). The research objective is to develop EST based SSR markers, in Cinnamon and to study its genes and other aspects. This plant is attribute with high medicinal value. It contains several secondary metabolites and other compounds like alkaloids, steroids. Flavonoids, saponins, proteins, tannins, polyphenols, and glycosides, known to impart specific

functions and hold important therapeutic roles.

Results: A total 7331 EST sequences of Cinnamon, were retrieved from the dbEST database in FASTA format. Among, these 1139 SSRs were identified, which includes 692 repeat unit for mononucleotides repeats, 161 repeat units for dinucleotide repeats, 270 repeat units for trinucleotide repeats, 11 repeat units for tetra nucleotide repeats, 1 repeat unit for penta-nucleotide repeats, and 4 repeat units for hexa-nucleotides repeats. Following identification of SSR, BLAST (Basic Local Alignment Search Tool) alignment of EST SSR was performed, which corroborated their functions. However, most of them were attributed to essential protein and many were gene related proteins, some of them were metabolically active proteins and enzymes were identified in the plant.

Conclusions: Findings will helps to analysis the important molecular markers and to facilitate the analysis of genetic diversity.

Keywords: Expressed sequence tags- simple sequence repeats, *in silico* studies, functional markers, cinnamon.

INTRODUCTION

Cinnamon or known as ‘true cinnamon’ is native to Sri Lanka and southern parts of the India. Cinnamaldehyde, eugenol, and linalool are the three main components of the essential oils obtained from the bark of Cinnamon, these components represent 82.5% of the total composition [1]. Trans cinnamaldehyde, accounts for approximately 49.9-62.8% of the total amount of bark oil [2,3]. Two more major components of cinnamon extracts are cinnamaldehyde and eugenol [4]. There are two main verities of cinnamon, *Cinnamomum zeylanicum* (CZ) and *Cinnamom cassia* (CC). These verities have basic difference in their coumarin (1,2-benzopyrone) content [5]. The levels of coumarins in CC seem to be very high and can cause health risk if consumed regularly in higher quantities. According to the German Federal Institute for Risk Assessment (BFR), 1 kg of CC powder contains approximately 2.1-4.4 g of coumarin, which means 1 teaspoon of CC powder would contain around 5.8-12.1 mg of coumarin. Above given is the TDI (Tolerable Daily Intake) for coumarin if 0.1 mg/kg body weight/day which was recommended by the European Food Safety Authority (EFSA) [6]. The BFR reports

precisely states that CZ contains ‘hardly any’ coumarin. Coumarins are secondary phyto-chemicals with strong anticoagulant, carcinogenic and having hepato-toxic properties [6]. The fundamental mechanisms for the coumarin content-related toxic effects are yet to be completely clarified. CC contains high concentration of coumarin than any other foods. Studies have shown that coumarin coverage from food consumption is mainly due to CC. Currently available evidences shown that coumarin does not appear to play any direct role in the observed biological effects of CC. However, CC variety has been shown many beneficial pharmaceutical properties [6,7]. Numerous beneficial health effects of CZ have been confirmed through *in-vitro* and *in-vivo* studies in animals. They have anti-inflammatory properties, reducing cardiovascular disease, anti-microbial activity, boosting cognitive function and reducing risk of colonic cancer. Cinnamon has been also mentioned in chinese texts as long as 4,000 years ago, it is one of the oldest herbal medicines known [8].

EST-SSR (expressed sequence tag-simple sequence repeat) is a new developed molecular marker based on the expression sequence of microsatellites. This technology has attained the advantage of avoiding the

construction steps of genomic DNA library in SSR development process; it gives the exact marker involve in gene function and shows similarity in genomic functional area. EST-SSR explains the phenotypic difference based on its polymorphisms. This EST-SSR are highly conserved within the species as it is a part of gene which leads to make the primers more commonly used among the species. Therefore for the development of SSR markers, these EST sequences act as valuable resources. In the recent years, several studies revealed that there are vast numbers of ESTs accumulated as the result of deep research analysis on different species. These accumulated EST data provides a platform in the development of SSR markers [9-11].

Various projects on sequencing or ESTs generates large amount of DNA sequence data which can be easily accessible to public, it carries both genic (EST) and genomic sequences which can be further used in the development of markers such as SSRs, SNPs. etc. The presence of any marker type from such data which can easily accessible leads to the generation of markers in cheap cost, like if SSRs are present in the genic sequence, they called as EST-SSRs [9].

The EST-SSR markers are associated with the genes carrying them as once they mapped. They also act as a valuable source of functional markers. Thus the formation of EST based SSR markers is a cheap alternative as compared to conventional SSR development method. In genome analysis of sorghum these EST-SSRs play a major role in producing lasting insight into processes by which novel genotypes are generated, such advantages helps in the applications of crop breeding programs [9-11].

The Conventional method of the development of SSR marker is tedious and costly. Therefore, the availability of genic EST sequence or genomic sequence in open public databases and availability of bioinformatics tools, the development of SSR marker is becoming now low cost and easier [12]. Although, previously several SSR markers were already generated by using EST databases in several crops. For the diversity analysis the EST-SSR markers were widely used in several crops like: wheat [10,13,14], barley [9], in mapping of barley [11, 15], pearl millet [16] and finger millet [17]. The genomic SSRs derived from the transcribed regions of the genome are more polymorphic as compare to the EST derived SSR markers [15, 12].

In the terms of cross-species transferability the EST-SSR markers are very superior, because they were derived from the most conserved regions of the genome which are very useful in the application of comparative genome mapping and phylogenetic analysis. EST-SSR markers developed in a small number (30) in sorghum with wheat, rice and maize [18]. These markers have also shown very transfer rate in several crops system. In wheat the EST-SSR markers developed showing 62% transferability across the all four species barley, maize, wheat and rice. EST-SSRs showing 40% transferability rate from barley to rice [11, 15].

MATERIALS AND METHODS

Development of EST-SSR markers: In the improvisation of species, molecular markers are prominently used, they help to identify the polymorphisms, mating system parameters, marker-assisted selection and genotype characterization. Finally EST-SSR was constructed for cinnamon as we found there was no EST-SSR developed till date.

Recognition of EST sequences

Firstly, the EST sequences of cinnamon were retrieved in FASTA format from the NCBI (<https://www.ncbi.nlm.nih.gov/>) i.e. National Centre for Biotechnology Information advances science & health by providing access to biomedical & genomic information. After that the MISA web (<http://webblast.ipk-gatersleben.de/misa/>) was used for the recognition and determination of the ideal microsatellite also compound microsatellites which are fitful by the certain number of bases from the ESTs recognized from the NCBI followed to design the primers at microsatellite loci.

MISA

In the plant genetics and the forensic science the microsatellites are prominently used marker system. The challenge is to make microsatellites from re-sequencing data. MISA is a web based computational application tool which help in the development of microsatellite markers. MISA web can be accessed by this link <http://misaweb.ipk-gatersleben.de/>. A 25 years ago microsatellites were rise and still it was a most common genetic marker using in plant breeding and plant genetics and forensics science, where it is generally known as simple sequence repeats (SSRs) or

short tandem repeats (STRs). In the microsatellite the basic structural block is the short sequence motifs present between one and six pairs in length which is repeated in tandem, by high throughput sequencing data or Sanger method these characteristics can be easily detected by giving *in-silico* approach using nucleotide sequences [10,12].

Pre-processing of the FASTA sequences

The retrieved FASTA sequence was pre-processed first by the help of software named CAP3 (<http://doua.prabi.fr/software/cap3>) which was freely available on web server, it identify the non-redundant EST sequences. The CAP3 software runs algorithm which overlaps between the sequences and further join the reads in the decreasing order to form contigs. After the pre-processing of FASTA sequences CAP3 gave two files ie. Contigs and Single tone which was further processed separately [12].

Selection of candidate EST sequences

The non-repeated SSR containing EST sequences of Cinnamon were used for homology search by using Basic Local Alignment Search Tool (BLAST) tool available in the NCBI. From all the BLAST

hits we identified an appropriate EST giving the maximum score was selected, followed by recognition of homologous genomic region. For the analysis of complete coverage across the genome sequence BLAST were performed.

Primer Designing

The selected contigs (SSR containing ESTs sequences) and the single tone were used to design primer pairs by using primer3 ([http://biotools.umassmed.edu/bioapps/prim er3_www.cgi](http://biotools.umassmed.edu/bioapps/primer3_www.cgi)). The Primers were designed in such a way that they follow such conditions: primer length (min-70nt, opt-160nt, max- 250nt), Tm (min-54°C, opt-57°C, max-60°C) & GC content (min-45%, opt-50%, max- 60%) [12].

BLAST

BLAST is a most common local alignment tool (Basic Local Alignment Search Tool) founded by Altschul. It is based on a set of algorithms in which a fragment of query sequence that aligns with the fragment of subject sequence present in the database. The initial alignment should be greater than neighbour score threshold (T). The alignments can be extended in both the direction till the score aligned segment is increase.

There were two alignments global and local. The global sequence approaches are used to compare the whole sequence with the other full sequences. In the local method the part of the sequence is align with the other part of the sequence. The global alignment gives comparison of one to other sequence, local alignment shows higher similarity in the regions but lack the ability of comparison of two sequences. While comparing small group of sequence global approach is very useful as the comparison of sequences increases the cost increases. The local alignments are based on heuristic programming approach that is very suitable for very large databases, but they do not provide give optimum solution. This limitation plays a major role in the genomics as they uncover regions of similarity that are correlated by two diverse sequences.

SWISS MODEL - ExPASy

ExPASy is the bioinformatics resource portal which gives a key to open scientific databases and software tools in other aspects of life sciences. It carries some useful tools like SWISS MODEL, UNIPROT, PROSITE and STRINGS WISS-MODEL it is a fully developed protein structure homology modelling server access by the ExPASy web

server. This server is used to make protein modelling accessible to the all researchers of life sciences worldwide; on the basis of FASTA sequence it provides the 3D structure of proteins.

RESULTS AND DISSCUSSION

MISA (Microsatellite Identification Search Tool)

The EST sequence which retrieve from the NCBI database, the CAP3, and MISA software is used for cinnamon plant, MISA gave the following results as discussed below:-

Distribution frequency of repeat units for all the SSR's in Cinnamon

RESULTS OF MICROSATELLITE SEARCH

=====
 Total number of sequences examined: 7
 Total size of examined sequences (bp): 5
 Total number of identified SSRs: 1
 Number of SSR containing sequences: 8
 Number of sequences containing more than 1 SSR: 1
 Number of SSRs present in compound formation: 2

Distribution to different repeat type classes

Unit size	Number of SSRs
1	692
2	161
3	270
4	11
5	1
6	4

Based on the results obtained from the cinnamon MISA analysis, In total 1139 SSRs were identified: out of which 692 repeat unit were mononucleotides repeats, 161 repeat unit for dinucleotide repeats, 270 repeat unit for trinucleotide repeats, 11 repeat unit for tetra nucleotide repeats, 1 repeat unit for penta-nucleotide repeats, and 4 repeat unit for hexa-nucleotides repeats.

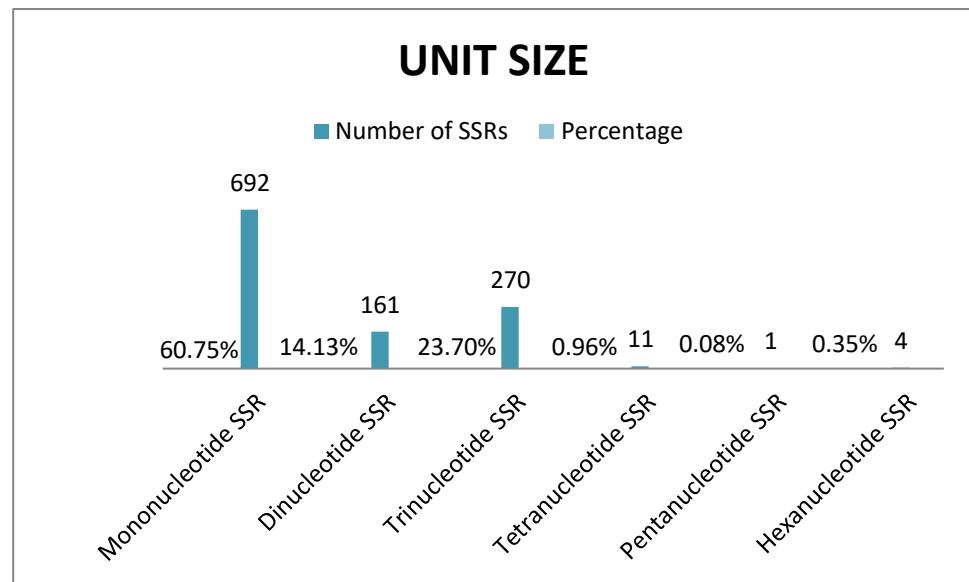


Figure1. Distribution frequency of repeat units for all the SSRs in Cinnamon.

Repeats	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
A/T	146	106	53	41	31	32	43	25	34	26	13	3	4	7	2	7
C/G	39	18	12	10	6	3	9	3	2	2	1					1
AC/GT	.	15	10	6	2	2						1									
AG/CT	.	24	5	4	2	6	1	1	3		6	1	4	1		2		1	1		
AT/AT	.	13	7	15	6	7	2		1		1	6	1					1			
AAC/GT																					
AAG/CTT	8	3	5																		
AAT/ATT	5	12	8	11	20	3															17
ACC/GGT	18	3	1																		
ACT/AGT	3		1																		
AGC/CTG	11	2	3	1																	
AGG/CTT	22	12			1																
ATC/ATG	54	2																			
CCG/GGG	11	26	4																		
AAAT/ATT	6	3																			
ACAT/ATG		1																			
AGGC/CTG	1																				
AAAG/CTTG	1																				
AAGCAG/CTGCTT																					3
ACCATG/ATGGTC		1																			

Figure 2: Distribution frequency of nucleotides for all the SSRs in Cinnamon.

The figure 2 indicated that A/T mononucleotide was the most common repeat among all SSR motifs, while the Most common dinucleotide motif was AG/CT. Among the tri nucleotide repeats AAT/ATT was most common; in tetranucleotide SSR motifs AAAT/ATT was most common. The pentanucleotide motif was AAAGC/CTTG, in hexanucleotide SSR motif

AAGCAG/CTGCTT was common. These results were similar to previous studies [12].

CAP3

Pre-processing of the EST sequences downloaded from the public domain were carried out by CAP3 software. By using the Cap3 program, which helps in the elimination of repeating data set from the sequence file, ultimately it results into the formation of two files with one containing contig sequence whereas in other the single tone sequence. As summary discussed below:-

Numbers of Contigs: 2233

Number of Single tone: 2305

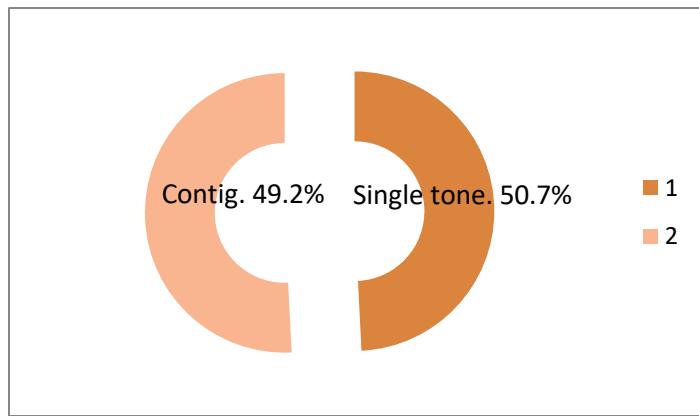


Figure 3: Distribution of single tone & contig in Cinnamon.

BLAST

BLAST was carried out by BLAST nucleotide analysis; all the 2233 Contig and 2305 Single tone sequences of Cinnamon were BLAST to analyze the putative function of the sequence. On the basis of their appropriate match, all SSR loci were divided into three groups;

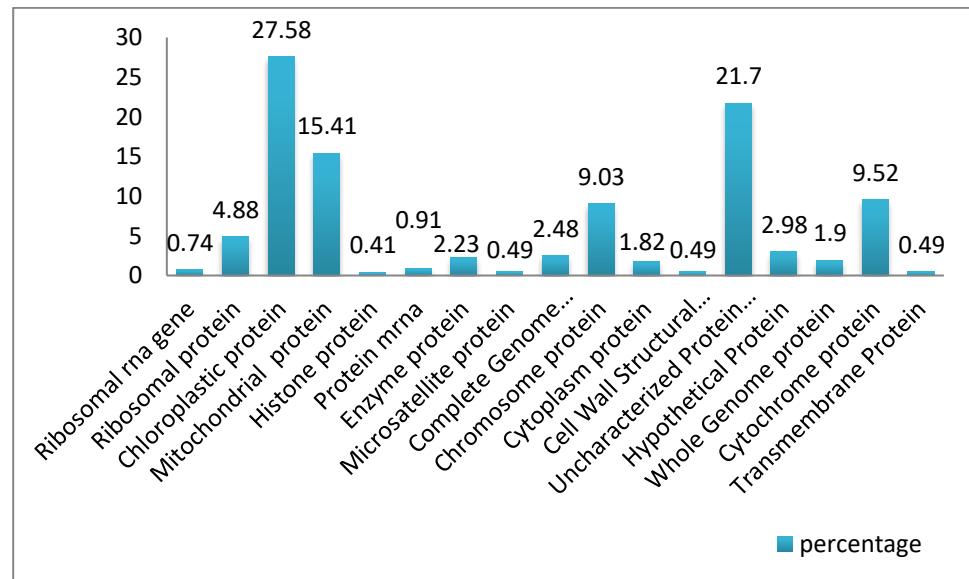


Figure 4: Biological distribution of contig and single tone.

Biological Function: In biological functions the genes acquire all vital processes like metabolism, photosynthesis, cell signalling, environmental related factors, etc. In this analysis we have got 1207 total sequences from Contig and Single tone from the Cap3 software there were found to be 9 Ribosomal RNA gene functionality, 59 Ribosomal Protein, 11 Protein mRNA, 333 chloroplastic & 186 mitochondrial proteins, 5 histone, 27 enzyme, 30 complete genome, 109 chromosome, 6 microsatellite, 22 cytoplasm, 6 cell wall structural, 262

uncharacterized protein, 36 hypothetical protein, 23 whole genome, 115 cytochrome & 6 transmembrane protein after the analysis.

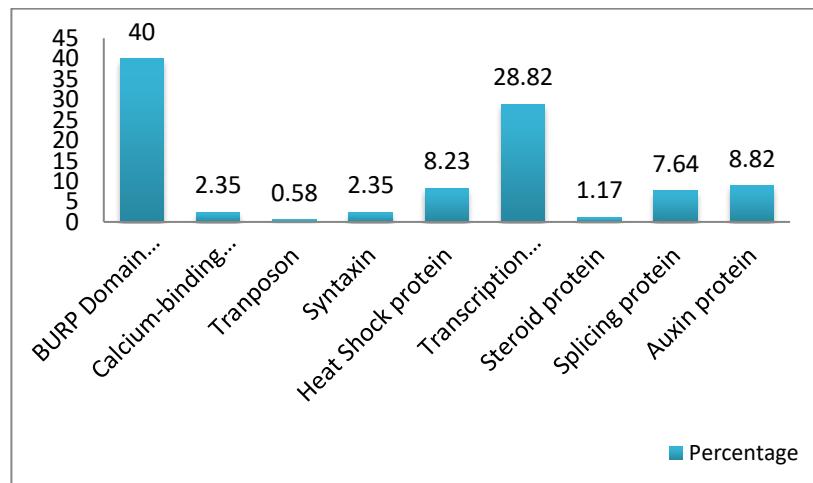


Figure 5: Protein distribution of contig and single tone.

Protein Function: By the help of nucleotide BLAST analysis, there were in total 170 proteins found in the sequences, among these 16 were heat shock proteins, 49 transcription proteins, 15 auxin, 2 steroid, 13 splicing, 68 BURP domain protein, 4 syntaxin, 4 calcium binding protein and 1 transposon.

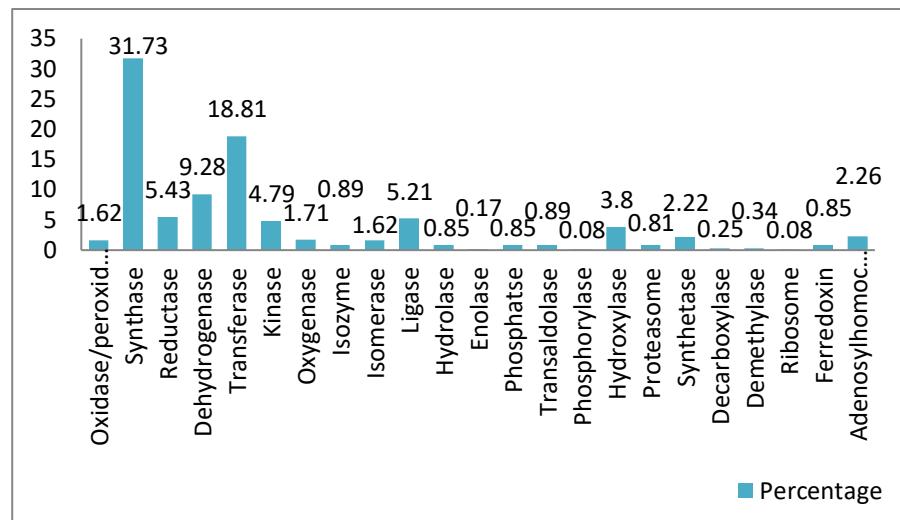


Figure 6: Enzyme distribution of contig and single tone.

Enzymatic Function: Cinnamon is important medicinal plant used for treating various diseases. It contains more than 2338 enzymatic activity with 742 synthases, 127 reductase, 38 oxidases & peroxidases, 21 isozymes, 38 isomerasers, 40 oxygenase, 217 dehydrogenase, 112 kinase, 565 transferases, 122 ligase, 20 hydrolase, 4 enolase, 20 phosphatase, 21 transaldolase, 2 phosphorylase, 89 hydroxylase, 19 proteasome, 52 synthetase, 6 decarboxylase, 8 demethylase, 2 ribosome, 20 ferredoxin and 53 adenosylhomocysteine.

Primer3

Primers were constructed based on suitable nucleotide and appropriate sequences after BLAST of contigs and single tone.

Table1: Characteristics of EST-derived SSRs for Cinnamon.

ID	TM (°C)	GC%	Forward Primer	Reverse Primer	Product size	Predicted function based on blast	Accession
DY327125.1	60	50	GCACCATCTTC GTCCTTCAT	TAACATTCCCC AGCTTCGTC	170	Erythrantheguttatus G-type lectin S-receptor-like serine/threonine-protein kinase At1g34300 (LOC105974395), mRNA	XM_012999490.1
DY327131.1	60	50	ATCCTCTGGAA GAGCTGCAA	TGATCAAGTG CGACCTTCAG	234	Erythrantheguttatus pyruvate kinase, cytosolic isozyme (LOC105956735), mRNA	XM_012980626.1
DY327152.1	60	53	ACTCATCTCGA CAGCCTCGT	CGGCACATCTT TCAGGAGAT	207	Agastache rugosa chalcone synthase (CHS) mRNA, complete cds	JQ314450.1
DY327176.1	60	45	CCTGGTTTTA ACGCTGGAA	GCCATGGGAT AGAGCAAAAA	249	Ocimumbasilicum germacrene D synthase (GDS) mRNA, complete cds	AY693644.1
DY327188.1	59	50	CGCACTCTTCA TCACTCCAA	ACTGCTATAA GCGCCATCGT	249	Sesamum indicum (RS)-norcoclaurine 6-O-methyltransferase-like (LOC105173010), mRNA	XM_011094646.2
DY327192.1	59	53	ACTGTTGGACC ATCCAGAGG	CCCAAAGCAA GAATCTCAGC	155	Sesamum indicum cyclin-dependent kinase D-3 (LOC105166737), transcript variant X2, mRNA	XM_020695628.1
DY327215.1	60	48	CCACTTCATGC TCCCTGTTT	GAAGCAAAAT TCGGTTGGAA	234	Sesamum indicum 3-phosphoshikimate 1-carboxyvinyltransferase 2 (LOC105171218), mRNA	XM_011092260.2
DY327278.1	60	50	ATGAGAAACA TGGCGAGGAC	TTCTTCTTCTC AGCGCCTTC	208	Sesamum indicum protein SRC1 (LOC105173924), mRNA	XM_011095843.2
DY327305.1	60	53	GAAGGACTTC	TGCTTAACAGC	162	Sesamum indicum serine/threonine-protein	XM_011088543.2

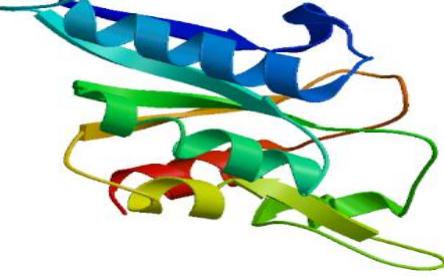
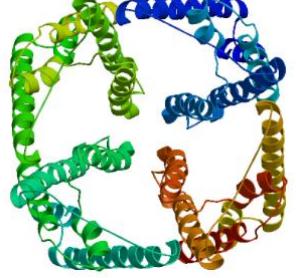
			CCCGATTCTC	AACGACCTG		kinase PBS1 (LOC105168454), mRNA	
DY327320.1	60	50	AGAGAGAGAT TCGCCGATCA	TTCGTCACTCG TGCTGAAAG	219	Olea europaea var. sylvestris serine/arginine-rich splicing factor SR45a-like (LOC111369703), transcript variant X3, misc_RNA	XR_002698229.1
DY327324.1	60	50	ATCCCATCCAT CCTTCCTTC	CGATCGACAC ATCGAAGCTA	155	Sesamum indicum glycosyltransferase family protein 64 protein C5 (LOC105174171), mRNA	XM_011096181.2
DY327360.1	60	50	AAACACAAGG TGCACCACAA	GCGATGGAGA GCCAACTTAG	180	Sesamum indicum autophagy-related protein 18f (LOC105176725), mRNA	XM_011099623.2
DY327460.1	60	50	CCTGGTTTA ACGCTGGAA	GCCATGGAT AGAGCAAAAA	249	Ocimumbasilicum germacrene D synthase (GDS) mRNA, complete cds	AY693644.1
DY327475.1	59	50	CAAGCTGTTCA ACCCAAAT	AGCGAGCTTC CTCATCTCAG	178	Sesamum indicum acyl-coenzyme A oxidase 3, peroxisomal (LOC105178460), mRNA	XM_011101927.2
DY327481.1	60	50	GCAAGGTAGT GCCAATCAT	GAAGTTGCGC AAGGCTAAC	177	Sesamum indicum 40S ribosomal protein S15 (LOC105171649), mRNA	XM_011092832.2
DY327482.1	59	55	ATCATTGTGG AGGGAGTGC	CCCTTGACCCC CTTAGACTC	199	Erythranthe guttatus serine hydroxymethyltransferase 4 (LOC105964521), transcript variant X2, mRNA	XM_012989028.1
DY327495.1	60	50	AGTGATCTCTT TGGGCATGG	TGAGAGCAAG GGAGGAGAAA	166	Ocimumbasilicum gamma-cadinene synthase (CDS) mRNA, complete cds	AY693645.1
DY327503.1	60	50	GAGGTCGAAG ATCCCCACAGA	TCAAATTGGTG CTCTTGCTG	176	Sesamum indicum serine hydroxymethyltransferase 4 (LOC105166533), mRNA	XM_011085916.2
DY327504.1	59	55	ATCATTGTGG AGGGAGTGC	CCCTTGACCC CTTAGACTC	199	Erythranthe guttatus serine hydroxymethyltransferase 4 (LOC105964521), transcript variant X2, mRNA	XM_012989028.1

SWISS MODEL-EXPASY

The 3D structures of important protein which were represented by contings and single tome sequences developed with the help of SWISS Model Expasy.

Table 2: Proteins Structures predicted on the basis of BLAST results.

PREDICTED: Sesamum indicum (RS)-norcoclaurine 6-O-methyltransferase-like (LOC105173010), mRNA	
PREDICTED: Sesamum indicumchorismate synthase 1, chloroplastic (LOC105166625), mRNA	

<p>PREDICTED: Erythranthe guttatus protein BPS1, chloroplastic-like (LOC105960562), transcript variant X2, mRNA</p>	
<p>PREDICTED: Sesamum indicum autophagy-related protein 18f (LOC105176725), mRNA</p>	
<p>PREDICTED: Sesamum indicum aquaporin TIP1-1-like (LOC105169946), mRNA</p>	

PREDICTED: *Sesamum indicum*
syntaxin-112-like (LOC105178362),
mRNA



CONCLUSION

In the present, developed EST-SSR will be highly useful in genotyping of cinnamon accessions with microsatellite markers, that can reveal the genetic diversity among accessions. These information will help us to select better parents with desired genes for the progeny to develop new commercial variety. It helps to generate novelty of the species with higher productivity and quality traits towards the sustainable development. The development of cinnamon SSR further helps in characterization of potential genetic makers which are very important for crop improvement and in gene mapping. These EST-SSR markers play a major role in determining the genetic relationship, pedigree analysis and genetic background of the species.

Acknowledgement: We wish to express our sincere acknowledgement to Dr. Ashok Kumar Chauhan, President, RBEF parent organization of Amity University Madhya Pradesh (AUMP), Dr. Aseem Chauhan, Additional President, RBEF and chairman of Amity University Gwalior Campus, Lt. Gen. V.K. Sharma, AVSM (Retd.), Vice Chancellor of AUMP Gwalior Campus, for

providing necessary facilities, their valuable support and encouragement throughout the work.

References:

- [1] Chericoni, S., Prieto, J.M., Iacopini, P., Cioni, P., & Morelli, I. (2005). In vitro activity of the essential oil of *cinnamomum zeylanicum* and eugenol in peroxynitrite induced oxidative processes. *Journal Agric Food Chem*, 53, 4762–4765.
- [2] Singh. G., Maurya, S., DeLampasona, M.P., & Catalan, C.A. (2007). A comparison of chemical, antioxidant and antimicrobial studies of cinnamon leaf and bark volatile oils, oleoresins and their constituents. *Food Chem Toxicol*, 45, 1650–1661.
- [3] Simic, A., Sokovic, M.D., Ristic, M., Grujic-Jovanovic, S., Vukojevic, J., & Marin, P.D. (2004). The chemical composition of some lauraceae essential oils and their antifungal activities. *Phytother Res*, 18, 713–717.
- [4] Usta, J., Kreydiyyeh, S., Barnabe, P., Bou-Moughlabay, Y., & Nakkash-Chmaisse, H. (2003). Comparative study on the effect of cinnamon and clove extracts and their main components on different types of ATPases. *Hum Exp Toxicol*, 22, 355–362.

[5] Archer, A. (1988). Determination of cinnamaldehyde, coumarin and cinnamyl alcohol in cinnamon and cassia by high-performance liquid chromatography. *J Chromatogr*, 447, 272–276.

[6] Abraham, K., Wöhrlin, F., Lindtner, O., Heinemeyer, G., Lampen, & A. (2010). Toxicology and risk assessment of coumarin: focus on human data. *Mol Nutr Food Res*, 54, 228–39.

[7] European Food Safety Association: Coumarin in flavourings and other food ingredients with flavouring properties. *EFSA J* 2008, 793:1–15.

[8] Verspohl, E.J., Bauer, K., & Neddermann, E. (2005). Antidiabetic effect of *Cinnamomum cassia* and *Cinnamomum zeylanicum* in vivo and in vitro. *Phytother Res*, 19(3):203-6.

[9] Cinnamon (*Cinnamomumzeylanicum*). Cinnamon - MotherNature.com Health Encyclopedia [Online] Available from: <http://www.mothernature.com/ency/Herb/Cinnamon.asp> (Accessed 2000 April).

[10] Varshney, R.K., Thiel, T., Stein, N., Langridge, P., & Graner, A. (2002). *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. *Cell Mol Biol Lett*, 7, 537-546.

[11] Gupta, P.K., Rustgi, S., Sharma, S., Singh, R., Kumar, N., & Balyan, H.S. (2003). Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Mol Gen Genom*, 270, 315-323.

[12] Mishra, R.K., Gangadhar, B.H., Nookaraju, A., Kumar, S., & Park, S.W. (2012). Development of EST-derived SSR markers in pea (*Pisum sativum*) and their potential utility for genetic mapping and transferability. *Plant breeding*, 131(1), 118-124.

[13] Leigh, F., Lea, V., Law, J., Wolters, P., Powell, W., & Donini, P. (2003). Assessment of EST- and genomic microsatellite markers for variety discrimination and genetic diversity studies in wheat. *Euphytica*, 133, 359-366.

[14] Zhang, L.Y., Ravel, C., Bernard, M., Balfourier, F., Leroy, P., Feuillet, C., & Sourdille, P. (2006). Transferable bread wheat EST-SSRs can be useful for phylogenetic studies among the Triticeae species. *Theor Appl Genet*, 113, 407-418.

[15] Thiel, T., Michalek, W., Varshney, R.K., & Graner, A. (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet*, 106, 411-422.

[16] Senthilvel, S., Jayashree, B., Mahalakshmi V., Kumar, P.S., Nakka, S.,

Nepolean, T., & Hash, C.T. (2008).

Development and mapping of simple sequence repeat markers for pearl millet from data mining of expressed sequence tags. *BMC Plant Biology*, 8,119.

[17] Dida, M.M., Srinivasachary, R.S., Bennetzen, J.L., Gale, M.D. & Devos, K.M. (2007). The genetic map of finger millet, *Eleusine coracana*. *Theoretic. Appl Genetics*, 114, 321-332.

[18] Temnykh, S., Park, W.D., Ayres, N., Cartinhour, S., Hauck, N., Lipovich, L., Cho, Y.G., Ishii, T., & McCouch, S.R. (2000). Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). *TheorAppl Genet*, 100,697-712.